基于多时间尺度数据的分布式光伏发电 预测模型

李启道¹,郝瑗宇¹,段鹏宇¹,赵 越¹,朱圆媛²

¹国网山西省电力有限公司,山西 太原 ²华北电力大学经济与管理学院,北京

收稿日期: 2024年10月28日; 录用日期: 2024年11月21日; 发布日期: 2024年11月28日

摘要

随着风电等新能源占比的不断提升,电力系统的气候依赖性日趋凸显。基于此,文章提出了一种基于多时间尺度数据的分布式光伏发电预测方法。构建多通道卷积神经网络,利用多通道CNN从多种时间尺度上提取影响因素数据所隐含的深层特征;借助BiLSTM神经网络解决序列信息长期依赖缺失问题;通过仿真实验,对比CNN-BiLSTM预测模型与基于多时间尺度数据的CNN-BiLSTM预测模型,验证新模型具有更高的预测精度。

关键词

多时间尺度,分布式光伏,CNN-BiLSTM预测

Distributed Photovoltaic Power Generation Prediction Model Based on Multi Time Scale Data

Qidao Li¹, Yuanyu Hao¹, Pengyu Duan¹, Yue Zhao¹, Yuanyuan Zhu²

¹State Grid Shanxi Electric Power Co., Ltd., Taiyuan Shanxi ²School of Economics and Management, North China Electric Power University, Beijing

Received: Oct. 28th, 2024; accepted: Nov. 21st, 2024; published: Nov. 28th, 2024

Abstract

With the continuous increase in the proportion of new energy such as wind power, the climate dependence of the power system is becoming increasingly prominent. Based on this, the article proposes a distributed photovoltaic power generation prediction method based on multi time scale data. The method builds a multi-channel convolutional neural network and uses multi-channel CNN to extract deep features hidden in influencing factor data from multiple time scales. Utilizing BiLSTM neural networks, the method also overcomes the long-term dependency issue in sequential data. Finally, the article compares the CNN BiLSTM prediction model with the CNN BiLSTM prediction model based on multi time scale data through simulation experiments and verifies that the new model has higher prediction accuracy.

Keywords

Multi Time Scale, Distributed Photovoltaics, CNN BiLSTM Prediction

Copyright © 2024 by author(s) and Hans Publishers Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

1. 引言

随着能源技术的进步和光伏发电理论的发展,分布式光伏发电正成为全球能源结构的重要组成部分。 光伏发电具有清洁、灵活性的优点,但同时其受多种天气因素影响,具有明显的波动性和间歇性特点。 提高光伏功率预测的准确性对于提升光伏发电的效率和稳定性是至关重要的。

光伏发电的短期预测目前可以分为两类。一类是基于物理模型的光伏发电系统短期发电预测方法, 另一类是基于光伏发电大数据的短期发电预测方法[1]。对于基于物理模型的光伏发电短期发电预测方法, 参考文献[2]通过确定光伏设备的开路电压、短路电流和两个形状参数,获得电流-电压输出曲线、填充 因子等的功率函数形式,然后获得光伏设备的输出功率。参考文献[3]提出了太阳能光伏电池的电压-电 流特性曲线和数学模型,解决了太阳能光伏电池工作环境变化多端、模型参数精度低的问题。仿真证明 了标准条件下模型输出的准确性。参考文献[4]提出了一种基于拉格朗日插值多项式的光伏电池 I-V 特性 建模方法,并以桑迪亚国家实验室的 5 个 I-V 特性曲线节点值作为拉格朗日插值的节点值,最终获得电 流-电压特性关系。上述光伏发电系统物理模型的短期功率预测方法主要通过模型化光伏设备的电流-电压特性曲线来获得最终输出功率。然而,上述方法在标准条件下的预测效果较好。但不确定环境温度 和光照强度等光伏设备所在环境因素的变化是否会影响最终结果。对于基于光伏发电大数据的短期功率 预测方法,主要分析气象数据、太阳辐射数据和光伏设备的历史输出功率等数据。通过大规模历史数据 训练,建立光伏发电功率与影响因素的适配关系,然后预测未来发电功率[5]。参考文献[6]采用结合数值 天气预测模型和机器学习的方法。将全球数据同化系统的气象数据输入人工神经网络(ANN)模型,对光 伏电站的发电功率进行拟合和预测。参考文献[7]提出了一种混合光伏发电功率预测模型。采用改进的前 向特征选择方法进行特征提取,解决了数据冗余和不完全特征提取等问题。模拟结果表明,该模型可以 有效提高光伏发电功率的预测精度和预测效率。参考文献[8]采用主成分分析降低多元数据序列的维度。 利用 LSTM 可以更好地反映时序数据的动态特性,提出一种基于 LSTM 网络的短期光伏发电功率预测方 法。参考文献[9]旨在解决光伏输出功率受多种因素影响、具有非线性和大波动的问题,并基于改进的深 度信念网络建立光伏短期功率预测模型。参考文献[10]提出了一种新的改进算法,以获得最佳的极限学习 机参数。改进的鸟群算法用于优化模型,以获得更好的预测结果。上述文献主要关注建立光伏输出功率 与输入变量之间适配关系的方法的研究。参考文献[11]结合了直接预测和间接预测。通过对预测日相似气 象条件下整数点的光伏发电功率进行统计分析,同时利用温度值、测量值和灰色模型的输出值构建神经 网络,最终获得准确的光伏发电功率预测值。参考文献[12]基于 XGBoost 建立了 XGBoost-LSTM 组合模 型。结合模型首先通过 XGBoost 模型进行初步预测,并将预测值作为新的输入来增加输入特征。然后通 过互反误差方法将 XGBoost 模型与 LSTM 模型结合起来,以提高预测精度。参考文献[13]针对光伏历史 数据缺失的问题,提出了一种基于二维序列填充框架和改进的 Kohonen 气象聚类的模型。采用二维序列 填充技术来弥补缺失数据,并考虑了影响光伏发电量的因素。参考文献[14]研究了在废弃光电限制条件下 的太阳能光伏电站的功率预测问题。提出了一种基于短期数值预测模型的模拟太阳辐射方法,并纠正了 辐射误差。在不同地区的光伏发电模拟中获得了良好的预测精度。参考文献[15]提出了一种基于径向基函 数神经网络的光伏发电功率预测模型。该模型使用模糊 K-均值算法对历史发电数据进行分类,并引入了 Harmony 搜索(HS)算法对其进行优化。对实际光伏发电系统的实验证明,该方法具有良好的预测质量。 参考文献[16]提出了一种基于变分模式分解和双重注意力机制 LSTM 的短期光伏功率预测方法。首先对 光伏功率序列进行分解,然后使用双重注意力机制 LSTM 对分解的组件进行预测。引入的注意力机制可 以自主提取关键时间信息点,更好地保证短期序列预测的稳定性。

2. 光伏发电输出功率影响因素特性分析

2.1. 发电输出功率影响因素

多种因素综合影响了光伏发电功率。这些因素既包括了气象因素、地理因素等外部因素的不确定性 冲击,也包括了电站设计、电气设备等内部因素的干扰,如图1所示。气象因素如全球水平辐射、气温、 大气压力、降雨量等变量变化波动较大,相比其他因素,对光伏发电输出功率的影响更为显著。地理因 素则由电站建设的位置、经纬度、海拔高度及周边遮挡条件等构成,由于电站建成后地理条件基本固定, 因而其对光伏发电输出功率的影响相比则较小。电站设计和电气设备也是影响光伏发电功率的重要因 素,其中前者包括组件安装角度和间距等因素,而后者则涉及电缆、变压器等设备的型号和转换效率 等因素。



Figure 1. Diagram of influencing factors on photovoltaic power generation 图 1. 光伏发电功率影响因素图

2.2. 发电输出功率影响因素相关性分析

多种内外部因素都可能会影响风电与光伏发电的输出功率,并且这些影响因素之间存在耦合关系, 彼此之间会相互影响。电站主要收集全球水平辐射、风速和风向、气温、降雨量等信息数据。各种因素 都会在不同程度上影响光伏电站的输出功率,同时其影响方式也各不相同,包括直接影响和间接影响。 预测模型中输入的每个变量都可能导致预测结果的变化,如果将现有的数据全部输入到模型中,庞大的 数据集势必会降低模型的运行效率。此外,若某一变量与光伏发电输出功率的关联性不明显,可能会导 致预测结果出现较大偏差。因此,为了适当缩减输入特征的输入维度,进而提高模型的预测准确性和运 算速度,本节将对各类影响因素进行相关性分析。

相关性分析基于统计学原理,通过计算相关性系数来明确各因素之间的关联程度和方向。相关性系 数是衡量因素重要性的关键指标,通过该系数可以区分主要因素和次要因素,最终将模型的输入变量限 制为主要因素对应的变量,通过减少低关联度的数据,提高数据的可靠性。

2.2.1. 皮尔逊相关系数

该系数作为衡量变量关联程度的指标,主要描述彼此间所具有的线性关系,该系数处在[-1,1]范围内。根据皮尔逊相关系数能够把变量关联程度划分为三个类别:正相关、负相关以及无关。系数表达式如下所示:

$$r_{p}(m,n) = \frac{\sum_{j=i}^{l} (m_{j} - \overline{m})(n_{j} - \overline{n})}{\sqrt{\sum_{j=i}^{l} (m_{j} - \overline{m})^{2}} \sqrt{\sqrt{\sum_{j=i}^{i} (n_{j} - \overline{n})^{2}}}$$
(1)

上式中, m代表光伏电站的历史出力; n代表输入预测模型的光伏发电影响因素。l代表样本个数, m_j 和 m分别代表电站的历史出力的第 j个值和平均值, n_j 和 n分别代表光伏发电的各类影响因素的第 j个值和平均值。皮尔逊相关系数等于变量和的协方差与其标准差之积的比值,表示 m和 n之间的线性关联程度。当 $r_p(m,n)$ 的绝对值越近似于 1,说明 m和 n之间的相关程度更高。如果 $r_p(m,n)=0$,说明此时 m 和 n之间无关联。

2.2.2. 灰色关联分析法

该方法在分析数据相关程度时首先需要构建参考序列与样本序列,其次需要完成对关联系数的计算, 再次需要计算灰色关联度,最后在确定关联度的基础上对其展开排序处理。具体算法流程如下:

1) 构建参考序列与比较序列:

$$U_0 = \{ U_0(1), U_0(2), \cdots, U_0(l) \}$$
⁽²⁾

$$U_{j} = \left\{ U_{j}(1), U_{j}(2), \cdots, U_{j}(l) \right\}$$
(3)

 U_0 表示参考序列指光伏电站出力的样本; U_j 表示比较序列矩阵指光伏电站的影响因素,系统有j个影响因素;影响因素以及出力的数据总量为l。

2) 计算比较序列与参考序列之间的关联系数:

$$\omega_{j}(l) = \frac{\min_{j} \min_{l} |U_{0}(l) - U_{j}(l)| + \psi \max_{j} \max_{l} |U_{0}(l) - U_{j}(l)|}{U_{0}(l) - U_{j}(l) + \psi \max_{i} \max_{l} |U_{0}(l) - U_{j}(l)|}$$
(4)

式中, **y**代表分辨系数。

3) 计算灰色关联度

为了便于比较相关性大小,必须对灰色关联度进行计算分析,灰色关联度是灰色相关系数的平均值。 最终将计算结果按照从大到小的顺序完成排序,得出不同影响因素与电站输出功率间的关联程度。

$$t_j = \frac{1}{n} \sum_{n=1}^{l=1} \psi_j(l) \tag{5}$$

由于灰色关联分析法通过计算数据关联系数的平均值求得数据的关联度,因此存在一定局限性。同时,两个序列之间可能存在负相关关系,灰色关联分析法对负关联度的讨论较少。考虑以上缺陷,本文采用皮尔逊相关系数法对光伏电站的历史影响因素(全球水平辐射、日射强度、天气温度、风速与风向、降雨量、雾霾浓度系数和光伏组件温度)进行特征选择,在保证预测结果准确性足够高的前提下,减少模型的输入影响因子以达到降低计算量的目的。



图 2. 皮尔逊相关分析热力图

从 Pearson 相关系数直方图中可以直观地看出,空气压力、风向、PM2.5 与光伏功率输出之间的相关 性极弱,而光伏模块温度、太阳辐射强度和全球水平辐射与光伏功率输出之间的相关性极高。为了减少 不利因素对模型预测准确性的影响,本文选取并保留影响因素数据集的 Pearson 相关系数大于 0.1 的数据 作为模型的输入变量。Pearson 相关性分析热力图如图 2 所示。虽然光伏功率输出与全球水平辐射和太阳 辐射强度之间的相关性都很高,但全球水平辐射和太阳辐射强度之间的相关性约为 1,这意味着全球水 平辐射和太阳辐射强度之间存在极强的相关性。因此,本文后续的计算中,仅保留全球水平辐射这一影响因素,而将全球水平辐射和太阳辐照强度的另一影响因素舍弃。本文使用 AQI 指数、PM2.5 和 PM10 三个指标来分析雾霾浓度与光伏发电功率之间的相关性。根据皮尔逊相关系数,在三个指标中,AQI 指数的相关性最高,PM2.5 的相关性最低。由于 AQI 包含 PM2.5 和 PM10,因此在本文后续的计算中,仅保留与光伏模块温度相关性最高的 AQI 作为影响因素。尽管光伏模块温度与气象温度、全球水平辐射和太阳辐照强度的相关性相对较高,但当太阳辐照强度增加时,相应的全球水平辐射和气象温度也会增加,此时光伏模块的温度也会增加。考虑到当光伏模块温度超过最佳工作温度时,太阳能电池板的光电转换效率会逐渐降低。本文后续的计算中,将光伏模块温度数据作为模型输入因素保留。然后,后续预测模型的输入数据为全球水平辐射、光伏模块温度、气象温度、AQI 和降雨量。

3. 基于多时间尺度数据的分布式电源发电预测模型

3.1. 模型构建

为了改善传统的 CNN 模型只能在单一时间尺度上提取数据特征导致潜在信息损失的问题,本节提出了一种基于多时间尺度数据的 CNN-BiLSTM 短期光伏发电输出功率预测模型,旨在对历史发电数据和影响因素数据进行不同时间间隔的处理,构建多通道卷积神经网络,分别提取不同时间尺度下的数据特征,从而增强特征多样性,避免有效信息的丢失。此外,引入鲸鱼优化算法,利用其自动优化和跳出局部最优的特性,对模型进行进一步优化。基于多时间尺度数据的 CNN-BiLSTM 模型结构如图 3 所示。



Figure 3. CNN-BiLSTM model structure for multi-timescale data 图 3. 多时间尺度数据的 CNN-BiLSTM 模型结构 模型以 CNN 网络模型为基础,构建了三个子模型。每个子模型将历史数据按照短间隔、短间隔和中间隔处理成输入数据,作为子模型的输入数据,传递到卷积层中。为了充分考虑不同时间尺度数据所包含的不同特征,本模型针对 1h、2h、3h 这三种时间尺度的数据分别建立了三个卷积通道,以提取不同时间尺度数据的特征信息。这种分别提取不同时间尺度的数据特征,并通过特征融合的方法将提取到的特征信息进行整合的方式能够确保输出特征既包含平滑后的信息,又保留原始输入特征,从而提升模型的预测性能。

首先由三个卷积核分别对1h、2h、3h 三个时间尺度的数据进行特征提取。之后添加池化层,并利用 Flatten 函数提升模型的泛化能力。随后,多通道 CNN 提取出的数据分别输入三个 Dropout 层,Dropout 层是神经网络训练中最常用也最有效的一种正则化方法,可以降低预测模型出现过拟合的现象。为了充分挖掘特征数据的时序关系,经 Dropout 层处理过的数据分别输送至三个 BiLSTM 网络中,使用 BiLSTM 网络提取时间特征,得到的输出数据继续通过 Dropout 层进行处理,输入 2 层全连接层,利用全连接层 对数据特征进行特征加权。最后利用 Concatenate 函数将三个不同时间尺度上的子模型进行特征融合,充分利用了不同时间尺度数据中包含的特征信息,并输出预测结果。

基于多时间尺度数据的 CNN-BiLSTM 短期光伏发电输出功率预测模型采用 CNN 神经网络进行特征 提取,使用 BiLSTM 神经网络进行预测,通过从多时间尺度数据集中提取不同发电特征以具备更强大的 特征提取能力,预测模型参数配置如表1所示。

超参数类型	超参数设置			
	1h时间尺度	2h时间尺度	3h时间尺度	
CNN 层数	1	1	1	
卷积核尺寸	2*2	2*2	2*2	
CNN 滤波器数	128	128	128	
BiLSTM 层数	1	1	1	
BiLSTM 层神经元数	128	128	128	

 Table 1. CNN-BiLSTM network hyperparameter configuration table based on multi-timescale data

 表 1. 基于多时间尺度数据的 CNN-BiLSTM 网络超参数配置表

3.2. 模型对比分析

本节选用 CNN-BiLSTM 模型与基于多时间尺度数据的 CNN-BiLSTM 模型进行对比分析。所选数 据集与上一节中 CNN-BiLSTM 模型所用数据集完全一致,均为山西省忻州 - 店坪站电站 2023 年 3 月 18 日至 3 月 30 日每 1 h 的数据集。图 4 为基于多时间尺度数据的 CNN-BiLSTM 模型输出功率对比图。 从图 4 中可以看出,绿色折线为 CNN-BiLSTM 模型预测功率曲线,橙色虚线为基于多时间尺度数据的 CNN-BiLSTM 模型预测功率曲线,蓝色实线为实际功率曲线,当输出功率曲线出现波峰值和波谷值时, 基于多时间尺度数据的 CNN-BiLSTM 模型的拟合表现更好,与新模型相比,CNN-BiLSTM 模型的预测 功率波动较大。因此,通过提取多时间尺度数据的特征信息,模型的预测精度有了进一步提升。

从表 2 的数据可以看出,本节所提出的基于多时间尺度数据的 CNN-BiLSTM 模型的 RMSE 值、MAE 值、MSE 值和 MAPE 值分别为 1.001、1.014、0.52 和 2.706。CNN-BiLSTM 模型的 RMSE 值、MAE 值、MSE 值和 MAPE 值分别为 1.033、1.028、0.71 和 3.475。相较于 CNN-BiLSTM 模型,基于多时间尺度数

据的 CNN-BiLSTM 模型的 RMSE 值降低了 3.10%、MAE 值减少了 1.36%、MSE 减少了 26.76%、MAPE 值减少了 22.13%。综合四个评价指标来看,基于多时间尺度数据的 CNN-BiLSTM 模型的预测性能更好,对于光伏发电输出功率的预测精准度有所提升。



Figure 4. Output power comparison of CNN-BiLSTM model based on multi-timescale data 图 4. 基于多时间尺度数据的 CNN-BiLSTM 模型输出功率对比

 Table 2. Evaluation metrics of CNN-BiLSTM model based on multi-timescale data

 表 2. 基于多时间尺度数据的 CNN-BiLSTM 模型评估指标

模型	RMSE	MAE	MSE	MAPE
CNN-BiLSTM	1.033	1.028	0.71	3.475
多时间尺度 CNN-BiLSTM	1.001	1.014	0.52	2.706

4. 总结

针对光伏发电系统的波动性、间歇性和复杂性等特点,本文提出基于多时间尺度数据的CNN-BiLSTM 短期分布式电源发电预测模型,通过截取山西省忻州-店坪站光伏电站的历史输出功率预测数据集与影响数据集的仿真分析,得出以下结论:

1) 所提出的基于多时间尺度数据的 CNN-BiLSTM 短期分布式电源发电预测模型拟合表现更好。当输出功率曲线出现波峰值和波谷值时,相较于 CNN-BiLSTM 模型,基于多时间尺度数据的 CNN-BiLSTM 模型预测功率波动较小,与实际功率曲线更为贴合。

2) 所提出的基于多时间尺度数据的 CNN-BiLSTM 短期分布式电源发电预测模型具有更好的预测性 能。相较于 CNN-BiLSTM 模型,基于多时间尺度数据的 CNN-BiLSTM 模型的 RMSE 值降低了 3.10%、 MAE 值减少了 1.36%、MSE 减少了 26.76%、MAPE 值减少了 22.13%,对于光伏发电输出功率的预测精 准度得到提升。

参考文献

- [1] 徐菁. 光伏发电系统短期功率预测方法的研究[D]: [硕士学位论文]. 南昌: 南昌大学, 2016.
- [2] Saetre, T.O., Midtgård, O. and Yordanov, G.H. (2011) A New Analytical Solar Cell I-V Curve Model. *Renewable Energy*, 36, 2171-2176. <u>https://doi.org/10.1016/j.renene.2011.01.012</u>
- [3] 孔令倩, 雷敏, 邓昭俊, 等. 基于 I-V 曲线的光伏电池特性仿真模型研究[J]. 新型工业化, 2014, 4(7): 28-34.

- [4] 郭婷婷, 伦淑娴. 基于拉格朗日插值多项式的光伏电池 I-V 特性建模方法[J]. 电子设计工程, 2015, 23(21): 145-147.
- [5] 宋智林. 光伏发电系统的短期发电功率预测研究[D]: [硕士学位论文]. 天津: 天津大学, 2016.
- [6] López Gómez, J., Ogando Martínez, A., Troncoso Pastoriza, F., Febrero Garrido, L., Granada Álvarez, E. and Orosa García, J.A. (2020) Photovoltaic Power Prediction Using Artificial Neural Networks and Numerical Weather Data. *Sustainability*, **12**, Article 10295. <u>https://doi.org/10.3390/su122410295</u>
- [7] Tao, C., Lu, J., Lang, J., Peng, X., Cheng, K. and Duan, S. (2021) Short-Term Forecasting of Photovoltaic Power Generation Based on Feature Selection and Bias Compensation-LSTM Network. *Energies*, 14, Article 3086. <u>https://doi.org/10.3390/en14113086</u>
- [8] 宋绍剑,李博涵. 基于 LSTM 网络的光伏发电功率短期预测方法的研究[J]. 可再生能源, 2021, 39(5): 594-602.
- [9] 梁彩霞, 高赵亮. 基于相似日和 GA-DBN 神经网络的光伏发电短期功率预测[J]. 电气应用, 2019, 38(3): 97-102.
- [10] 饶宇飞, 刘阳, 李玲玲, 等. 基于改进鸟群算法和极限学习机模型的光伏发电系统输出功率预测研究[J]. 可再生 能源, 2020, 38(10): 1318-1325.
- [11] 王守, 张娜. 基于灰色神经网络组合模型的光伏短期出力预测[J]. 电力系统自动化, 2012, 36(19): 37-41.
- [12] 谭海旺,杨启亮,邢建春,等.基于 XGBoost-LSTM 组合模型的光伏发电功率预测[J].太阳能学报,2022,43(8): 75-81.
- [13] 李燕青, 杜莹莹. 基于双维度顺序填补框架与改进 Kohonen 天气聚类的光伏发电短期预测[J]. 电力自动化设备, 2019, 39(1): 60-65.
- [14] 崔杨,陈正洪,孙朋杰.弃光限电条件下不同纬度地区短期光伏发电功率预测对比分析[J].太阳能学报,2018, 39(6):1610-1618.
- [15] Huang, C., Chen, S., Yang, S. and Kuo, C. (2015) One-Day-Ahead Hourly Forecasting for Photovoltaic Power Generation Using an Intelligent Method with Weather-Based Forecasting Models. *IET Generation, Transmission & Distribution*, 9, 1874-1882. <u>https://doi.org/10.1049/iet-gtd.2015.0175</u>
- [16] 杨晶显, 张帅, 刘继春, 等. 基于 VMD 和双重注意力机制 LSTM 的短期光伏功率预测[J]. 电力系统自动化, 2021, 45(3): 174-182.